

NOTES FROM THE FIELD

Buhi'non (Bikol) Digital Wordlist: Presentation Form

Kenneth S. Olson

SIL International and University of North Dakota

Emy T. Ballenas

Translators Association of the Philippines

Nilo M. Borromeo

Translators Association of the Philippines

This paper presents a 207-item digital wordlist of Buhi'non, an Austronesian language spoken in the Philippines. The wordlist includes a broad phonetic transcription of each word, an English gloss, an individual WAV recording of each item, and metadata for resource discovery. An archival form of the wordlist was deposited into an institutional archive (Kaipuleohone, the University of Hawai'i Digital Ethnographic Archive) and includes the original WAV file of the entire wordlist recording session, descriptive markup encoding of the wordlist in XML employing Unicode transcription, TIFF images of the original field transcriptions, and the metadata record. The presentation form was then generated directly from the archival form.

1. INTRODUCTION.¹ This paper presents a 207-item digital wordlist of core vocabulary in Buhi'non (ISO 639–3 code *bhk*, Lewis 2009), a variety of Southern (Inland) Bikol spoken in the town of Buhi, Camarines Sur Province, Philippines (cf. McFarland 1974, Lobel 2006). This presentation form was generated from an archival form of the data. The procedure we followed for creating both forms is detailed in Simons et al. 2007 and summarized below. Previous descriptive wordlists of Buhi'non are found in Yamada 1972 and McFarland 1974:306–312.

In addition to a description of the primary data in the form of phonetic transcription, we provide a documentation of the data in the form of digital audio recordings (cf. Himmelmann 1998, Himmelmann 2006:17–25), enabling the reader to verify and critique our transcription. Of interest in Buhi'non is the occurrence of an extra consonant phoneme

¹We wish to thank Josephine Daguman, Gary DuBois, Judy Kuntz, Alfred Mijares, Racquel Mijares, Roger Olson, Steve Quakenbush, Dean Salter, Gary Simons, Translators Association of the Philippines, and SIL-Philippines for technical and logistic assistance; two anonymous reviewers for helpful comments; and SIL Asia Area for funding. The data for this paper were collected during a Phonetics and Phonology Workshop organized by the Linguistics and Ethnology Department of the Translators Association of the Philippines, April 2006.

/y/ and an extra vowel phoneme /ə/ with respect to most other Bikol varieties (cf. Lobel 2006:771).

The materials included in this presentation of the data include the following:

- Wordlist: the present report, containing for each word an English gloss, a broad phonetic transcription, and pertinent notes.
- Recordings: digital recordings of each item are accessible by clicking on the loudspeaker icon next to the phonetic form of each word in the list below. Macintosh users need Acrobat reader to play embedded sounds; they cannot be played in Macintosh Preview. WAV² files of each item on the list are available at <http://hdl.handle.net/10125/4440>.
- Metadata: a description of the data, useful for resource discovery.

The original wordlist materials included two items: a six-page paper wordlist and a 13-minute digital recording in WAV format. The commonly used wordlist questionnaire that we employed presented 207 items; this is an amalgamation of the Swadesh 200 (Swadesh 1952:456–457) and Swadesh 100 (Swadesh 1955:124, 133–137) wordlists, with some minor modifications (*thou* → *you sg.*, *ye* → *you pl.*, *person* → *man (human being)*, *woods* → *forest*, *berry* → *fruit*, *claw* → *finger nail*, and *right* → *correct*). For each item, the questionnaire provided a prompt in English and space for the transcription of the elicited form. The second and third authors worked together to translate the wordlist into Buhi'non.

The wordlist was recorded on April 14, 2006 in the recording studio at the SIL Center in Bagabag, Nueva Vizcaya Province, Philippines. During the recording session, the second author produced the English prompt, and the third author produced the target word twice. The recording was made using a Samson C01U USB studio condenser microphone mounted on a table-top microphone stand and connected to a notebook computer, using Speech Analyzer v. 2.7 for audio capture.

The third author is a 23-year-old male native speaker of Buhi'non from the town of Buhi. Both of his parents are native Buhi'non speakers. He has received formal education through college, but has lived in Buhi all of his life. Besides Buhi'non, he also speaks Tagalog, English, Iriga Bikol, and Standard Bikol.

The microphone and software allowed for recording at a sampling rate of 44.1 kHz and a quantization of 16-bit (i.e., standard audio CD quality). This is sufficient for technical purposes since it covers nearly all acoustic information pertinent to language. The recordings were stored in non-compressed WAV format, the recommended industry standard (IASA-TC03 2005:8). The six-page paper wordlist was digitized on April 17, 2006 by the first author on an Epson Perfection 1260 scanner at archival quality (8-bit grayscale, 300 dpi, saved as uncompressed TIFF files, cf. MATRIX 2001).

² Definitions of WAV, XML, XSLT, and other pertinent terms employed here are available at <http://emeld.org/school/glossary.html>.

The draft transcription contained fields for three annotations: the item number, an English prompt, and a broad phonetic transcription of the Buhi'non response. The authors modified the phonetic transcription to make it conform to the International Phonetic Alphabet (IPA 1999), and added clarifying notes to certain entries. These data were entered into a Comma Separated Values (CSV) file using Unicode characters (<http://www.unicode.org>) and imported into TableTrans v. 1.2 software (Bird et al. 2002), where they were time-aligned to the original audio recording by Judy Kuntz at the International Linguistics Center in Dallas, Texas in December 2007. This annotation was exported to an XML annotation graph output (<http://www.w3.org/XML/>) and transformed into an XML descriptive wordlist format using an XSLT script.

The original WAV file, the digitized TIFF field transcriptions, the XML descriptive wordlist, and a metadata record constitute the archival form of the wordlist (Ballenas et al. 2009). The metadata record follows the standard set up by the Open Language Archives Community (<http://www.language-archives.org/OLAC/metadata.html>).³ The archival materials can be downloaded from Kaipuleohone, the University of Hawai'i Digital Ethnographic Archive (Albarillo and Thieberger 2009). The presentation form of the wordlist was then generated from the archival form. An XSLT script was employed to convert the archival XML descriptive wordlist into an HTML presentation wordlist. Then, TableTrans was used to automatically create individual sound files corresponding to each of the segments identified in the transcription process for use in the presentation form.

The data in this wordlist include certain limitations. First, the items are spoken in isolation rather than in carrier sentences. A carrier sentence might help avoid list intonation and make it easier to measure the length of some sounds (e.g., word-initial stops) (cf. Ladefoged 2003:7–8).⁴ Second, only one speaker was recorded. Recording a larger number of speakers could help ensure that the data reflect the language as a whole and not an individual's idiolect. Ladefoged (2003:14) recommends recording half a dozen speakers of each sex. Third, because we lacked the necessary hardware, the recordings do not meet the generally-accepted recommended best practice of 96 kHz and 24-bit for archival data (IASA-TC03 2005:8; Ladefoged 2003:18, 26; Plichta and Kornbluh 2002; Simons et al. 2007:31). We recommend that field researchers record primary data at archival-quality rates if possible. Fourth, some tokens exhibit background noise which may have been due to user contact with the microphone stand.


















On the other hand, the project succeeded in following most of the recommendations for best practice found in Bird and Simons 2003:574–579 and Simons et al. 2007. For example, the entire primary recording and original field transcriptions were stored in open

³ An alternative standard for linguistic metadata has been proposed called the ISLE Metadata Initiative. See <http://www.mpi.nl/IMDI/> for more details.

⁴ A reviewer points out that the use of carrier sentences can also be problematic. For example, word-final glottal stops in Tagalog tend to be elided except in phrase-final position. A carrier sentence would mask their presence.

(i.e., published and nonproprietary) formats. The transcriptions were made in IPA, encoded in Unicode, time-aligned to the underlying recording, and stored in an open descriptive markup format. The archival form was deposited into an institutional archive dedicated to the long-term preservation and availability of the resource. A human-readable version of the resource using presentational markup was provided (i.e., this web resource). The resource was described using metadata listed with an industry-standard repository for resource discovery. Most important, the present resource makes available for the first time primary data in the form of audio recordings for this under-researched language.

2. WORDLIST.

No.	English	Phonetic	Notes
1.	I	 [ʔa'ko]	
2.	you sing.; thou	 [ʔi'ka]	
3.	he	 [ʔi'ja]	
4.	we	 [ki'ta]	
5.	you pl.	 [ka'mo]	
6.	they	 [si'ra]	
7.	this	 [ʔamwa'di]	literally 'It is this (near the speaker)'; cf. [ʔa'di] 'this'
8.	that	 [ʔamwa'ʔan]	literally 'It is that (near the addressee)'; cf. [ʔa'ʔan] 'that'
9.	here	 [di'di]	
10.	there	 [ʔad'to]	far from both speaker and addressee
11.	who	 [siʔi'saj]	
12.	what	 [ʔo'no]	
13.	where	 [saʔ'ri]	
14.	when	 [kinao'no]	
15.	how	 [pao'no]	
16.	not	 [ʔin'diʔ]	
17.	all	 [ŋaʔ'min]	

18.	many	▶▶ [mari'gən]	intensifier; 'very' when accompanying an adjective; 'many' when accompanying a count noun
19.	some	▶▶ [na,ŋag,kapi'ra]	root: [pi'ra] 'how many'
20.	few	▶▶ [ʔag'baɟ]	
21.	other	▶▶ [ʔi'ba]	
22.	one	▶▶ [ʔə'sad]	speaker produced [ʔi'sad]
23.	two	▶▶ [do'wa]	
24.	three	▶▶ [to'lo]	
25.	four	▶▶ [ʔə'pat]	
26.	five	▶▶ [li'ma]	
27.	big	▶▶ [dakə'ʔəʔ]	
28.	long	▶▶ [ʔala'baʔ]	
29.	wide	▶▶ [maʔi'was]	
30.	thick	▶▶ [maiʔ'bəɟ]	
31.	heavy	▶▶ [mabəʔ'gat]	
32.	small	▶▶ [sa'daɟ]	
33.	short	▶▶ [ʔaliʔ'pot]	as in length; e.g., short rope; short message
34.	narrow	▶▶ [pigi'pit]	
35.	thin	▶▶ [mani'pis]	as in thickness; e.g. a thin book; thin hair
36.	woman	▶▶ [baba'ji]	
37.	man (adult male)	▶▶ [laɣa'ki]	
38.	man (human being)	▶▶ [ta'wo]	
39.	child (a youth)	▶▶ [ʔa'gin]	
40.	wife	▶▶ [sə'ʔəd na baba'ji]	literally 'woman spouse'
41.	husband	▶▶ [sə'ʔəd na laɣa'ki]	literally 'man spouse'
42.	mother	▶▶ [ʔi'naʔ]	
43.	father	▶▶ [ʔa'maʔ]	
44.	animal	▶▶ [ʔa'jəp]	
45.	fish	▶▶ [ʔis'raʔ]	
46.	bird	▶▶ [ba'joŋ]	
47.	dog	▶▶ [ʔa'jam]	

48.	louse	🔊	[ko'to]	
49.	snake	🔊	[ʔa'ʏas]	
50.	worm	🔊	[ʔo'ʏod]	
51.	tree	🔊	[ka'woj]	
52.	forest	🔊	[kakawo'jan]	root: [ka'woj]; alternate form: [kakaw'jan]
53.	stick (of wood)	🔊	[gi'riw]	
54.	fruit	🔊	[pro'tas]	loan
55.	seed	🔊	[toʔ'wan]	
56.	leaf	🔊	[da'won]	
57.	root	🔊	[ga'mot]	
58.	bark (of tree)	🔊	[ʔo'bak]	
59.	flower	🔊	[bo'rak]	
60.	grass	🔊	[ʔa'wot]	
61.	rope	🔊	[lo'bid]	
62.	skin (of a person)	🔊	[ko'lit]	
63.	meat (as in flesh)	🔊	[la'man]	
64.	blood	🔊	[ro'goʔ]	
65.	bone	🔊	[toʔ'wan]	
66.	fat (n.)	🔊	[ta'baʔ]	
67.	egg	🔊	[ʔo'mon]	
68.	horn	🔊	[so'ŋoj]	
69.	tail	🔊	[ʔi'kog]	
70.	feather	🔊	[ga'bi]	larger feathers rather than down
71.	hair	🔊	[bo'wok]	
72.	head	🔊	[pa'jo]	
73.	ear	🔊	[tali'ŋa]	
74.	eye	🔊	[ma'ta]	
75.	nose	🔊	[ʔo'ron]	
76.	mouth	🔊	[ŋo'soʔ]	
77.	tooth	🔊	[ŋi'pən]	front rather than molar
78.	tongue	🔊	[ri'raʔ]	

79.	fingernail	🔊	[ko'ko]	
80.	foot	🔊	[pa'ʔa]	
81.	leg	🔊	[pa'ʔa]	
82.	knee	🔊	[to'wod]	
83.	hand	🔊	[ka'mət]	
84.	wing	🔊	[pak'pak]	
85.	belly	🔊	[tə'jan]	
86.	guts	🔊	[bito'ka]	
87.	neck	🔊	[li'ʔəg]	
88.	back	🔊	[taɣod'tod]	
89.	breast	🔊	[ɣəɣəd'gan]	literally 'chest'; euphemism for 'breast'; cf. [so'so] 'breast'
90.	heart	🔊	[po'soʔ]	
91.	liver	🔊	[ʔa'taj]	
92.	drink	🔊	[ʔi'nom]	
93.	eat	🔊	[ʔa'boʔ]	'angry' form; standard form:[ka'ʔən]
94.	bite	🔊	[ʔəb'ʔəb]	
95.	suck	🔊	[səp'səp]	
96.	spit	🔊	[lo'daʔ]	
97.	vomit	🔊	[so'ka]	
98.	blow (as wind)	🔊	[pa'rəs]	
99.	breathe	🔊	[niʔi'ŋa]	
100.	laugh	🔊	[ŋak'ŋak]	
101.	see	🔊	[sə'ɣəŋ]	'look' or 'watch'
102.	hear	🔊	[rə'ŋəg]	speaker produced initial [s]
103.	know (a fact)	🔊	[ʔi'si]	
104.	think	🔊	[ʔi'sip]	
105.	smell (sense odor)	🔊	[pa'roŋ]	
106.	fear	🔊	[ta'kot]	
107.	sleep	🔊	[to'rog]	
108.	live	🔊	[bwaj]	
109.	die	🔊	[gəra'ʔan]	
110.	kill	🔊	[gəraʔ'nən]	
111.	fight	🔊	[la'ban]	

112.	hunt (transitive)	▶▶	[raʔ'kəp]	
113.	hit	▶▶	[so'gat]	as in target; e.g. rifle or boxing
114.	cut	▶▶	[pot'lon]	root: [po'tol]; synonym: [ta'ras]
115.	split	▶▶	[ba'ʔak]	e.g., fruit; cf., [ti'pak] 'split (e.g. bamboo or rock)'
116.	stab (or stick)	▶▶	[bo'noʔ]	
117.	scratch (an itch)	▶▶	[ka'gaw]	
118.	dig	▶▶	[nika'wot]	
119.	swim	▶▶	[la'ŋoj]	
120.	fly (v.)	▶▶	[la'jog]	
121.	walk	▶▶	[ʔa'gi]	
122.	come	▶▶	[ʔa'bot]	
123.	lie (as on one's side)	▶▶	[saŋ'dəj]	literally 'to lie down with feet on top of someone else who is also lying down'
124.	sit	▶▶	[ʔə'ʔaʔ]	
125.	stand	▶▶	[tin'dəg]	
126.	turn (change direction)	▶▶	[bi'rik]	
127.	fall (as in drop)	▶▶	[ʔo'log]	
128.	give	▶▶	[ta'ʔo]	
129.	hold (in one's hand)	▶▶	[kap'tan]	
130.	squeeze	▶▶	[pə'gaʔ]	
131.	rub	▶▶	[kis'kis]	
132.	wash	▶▶	[ʔo'gas]	
133.	wipe	▶▶	[po'nas]	
134.	pull	▶▶	[bə'təŋ]	
135.	push	▶▶	[tod'loŋ]	
136.	throw	▶▶	[ʔə'mət]	
137.	tie	▶▶	[gəʔ'ran]	
138.	sew	▶▶	[ta'jiʔ]	
139.	count	▶▶	[bi'laŋ]	
140.	say	▶▶	[sab'jən]	

141.	sing	▶▶	[kan'ta]	loan
142.	play	▶▶	[ja'mən]	
143.	float	▶▶	[lə'taw]	
144.	flow	▶▶	[bo'los]	
145.	freeze	▶▶		no Buhi'non equivalent
146.	swell	▶▶	[gə'tək]	
147.	sun	▶▶	[ʔal'daw]	
148.	moon	▶▶	[bo'lan]	
149.	star	▶▶	[bito'ʔon]	
150.	water	▶▶	[to'big]	
151.	rain	▶▶	[ʔo'ran]	
152.	river	▶▶	[sa'log]	
153.	lake	▶▶	[la'wəd]	
154.	sea (as in ocean)	▶▶	[da'gat]	
155.	salt	▶▶	[ʔa'sin]	
156.	stone	▶▶	[ba'to]	
157.	sand	▶▶	[baj'baj]	
158.	dust	▶▶	[ʔal'pog]	
159.	earth (as in soil)	▶▶	[ra'gaʔ]	
160.	cloud	▶▶	[paŋano'ron]	
161.	fog	▶▶	[ʔam'bon]	
162.	sky	▶▶	[la'ŋit]	
163.	wind (as in breeze)	▶▶	[pa'rəs]	
164.	snow	▶▶		no Buhi'non equivalent
165.	ice	▶▶	[ji'lo]	
166.	smoke	▶▶	[ʔa'so]	
167.	fire	▶▶	[kala'jo]	
168.	ashes	▶▶	[ʔa'bo]	
169.	burn (intransitive)	▶▶	[so'loʔ]	
170.	road	▶▶	[tinam'po]	
171.	mountain	▶▶	[bo'kid]	
172.	red	▶▶	[po'la]	
173.	green	▶▶	[bir'di]	loan

174.	yellow	🔊	[ji'lo]	loan
175.	white	🔊	[po'tiʔ]	
176.	black	🔊	[ʔi'təm]	
177.	night	🔊	[gaʔ'bi]	
178.	day (daytime)	🔊	[ʔal'daw]	
179.	year	🔊	[ta'ʔon]	
180.	warm (as in weather)	🔊	[laʔi'nit]	'very hot (water or weather)'
181.	cold (as in weather)	🔊	[mali'pət]	
182.	full	🔊	[po'noʔ]	
183.	new	🔊	[ba'go]	
184.	old	🔊	[go'raŋ]	
185.	good	🔊	[ma'raj]	
186.	bad	🔊	[mara'ʔət]	
187.	rotten (as a log)	🔊	[lə'paʔ]	
188.	dirty	🔊	[maʔa'tiʔ]	
189.	straight	🔊	[to'lid]	loan from Iriga Bikol; the Buhi'non word is [tog'joʔ]
190.	round	🔊	[bi'log]	
191.	sharp (as a knife)	🔊	[mata'rəm]	
192.	dull (as a knife)	🔊	[maŋə'rəw]	
193.	smooth	🔊	[maki'nis]	
194.	wet	🔊	[sə'məg]	
195.	dry (adjective)	🔊	[ma'ra]	
196.	correct (right)	🔊	[taɣ'ma]	variant of the standard form [ta'ma]
197.	near	🔊	[ʔara'ni]	
198.	far	🔊	[ʔara'joʔ]	
199.	right (side)	🔊	[to'ʔo]	

- | | | | | |
|------|---------------------|---|--------------|---|
| 200. | left (side) |  | [wa'la] | |
| 201. | at |  | [ʔadto 'sa] | literally 'there at/in/on';
cf. item 202 |
| 202. | in |  | [sa] | also 'on'; 'at'; 'into'; etc. |
| 203. | with (accompanying) |  | [kaʔiba'nan] | |
| 204. | and |  | [sag'kəd] | |
| 205. | if |  | [kin] | |
| 206. | because |  | [da'il] | |
| 207. | name |  | [ŋa'ran] | |

REFERENCES

- ALBARILLO, EMILY E., and NICK THIEBERGER. 2009. Kaipuleohone, the University of Hawai'i's digital ethnographic archive. *Language Documentation & Conservation* 3(1):1–14. <http://hdl.handle.net/10125/4422>.
- BALLENAS, EMY T., NILO M. BORROMEO, KENNETH S. OLSON, JUDY KUNTZ, ROGER E. OLSON, and GARY F. SIMONS. 2009. Buhi'non (Bikol) digital wordlist: Archival form. Kaipuleohone, the University of Hawai'i Digital Ethnographic Archive, Honolulu, HI. <http://hdl.handle.net/10125/4916>.
- BIRD, STEVEN, KAZUAKI MAEDA, XIAOYI MA, HAEJOONG LEE, BETH RANDALL, and SALIM ZAYAT. 2002. TableTrans, MultiTrans, InterTrans and TreeTrans: Diverse tools built on the Annotation Graph Toolkit. *Proceedings of the Third International Conference on Language Resources and Evaluation*. Paris: European Language Resources Association. <http://arxiv.org/abs/cs/0204006>.
- BIRD, STEVEN, and GARY SIMONS. 2003. Seven dimensions of portability for language documentation and description. *Language* 79(3):557–82.
- HIMMELMANN, NIKOLAUS. 1998. Documentary and descriptive linguistics. *Linguistics* 36: 161–95.
- HIMMELMANN, NIKOLAUS. 2006. Language documentation: What is it and what is it good for? In *Essentials of language documentation*, ed. by Jost Gippert, Nikolaus P. Himmelmann, and Ulrike Mosel 1–30. Berlin: Mouton de Gruyter.
- IASA-TC03. 2005. The safeguarding of the audio heritage: Ethics, principles and preservation strategy, version 3. http://www.iasa-web.org/downloads/publications/TC03_English.pdf.
- INTERNATIONAL PHONETIC ASSOCIATION. 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.
- LADEFOGED, PETER. 2003. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Oxford: Blackwell.
- LEWIS, M. PAUL, ed. 2009. *Ethnologue: Languages of the world*, 16th edition. Dallas, TX: SIL International. <http://www.ethnologue.com>.
- LOBEL, JASON W. 2006. Bikol. In *Encyclopedia of language & linguistics*, 2nd edition, ed. by Keith Brown, 769–72. Amsterdam: Elsevier.
- MATRIX: THE CENTER FOR HUMANE ARTS, LETTERS AND SOCIAL SCIENCES ONLINE AT MICHIGAN STATE UNIVERSITY. 2001. Digital imaging for archival preservation and online presentation: Best practices. ms. http://www.historicalvoices.org/papers/image_digitization2.pdf.
- McFARLAND, CURTIS DANIEL. 1974. The dialects of the Bikol area. PhD dissertation, Yale University.
- PLICHTA, BARTEK, and MARK KORNBLUH. 2002. Digitizing speech recordings for archival purposes. http://www.historicalvoices.org/papers/audio_digitization.pdf.
- SIMONS, GARY F., KENNETH S. OLSON, and PAUL FRANK. 2007. Ngbugu digital wordlist: A test case for best practices in archiving and presenting language documentation. *Linguistic Discovery* 5(1):28–39. <http://journals.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/2/xmlpage/1/article/314>.

- SWADESH, MORRIS. 1952. Lexico-statistic dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96(4):452–63.
- SWADESH, MORRIS. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21:121–37.
- YAMADA, YUKIHIRO. 1972. *A Buhi Bikol vocabulary*. Bulletin of the Faculty of Education, Kochi University, No. 24, Series I. Kochi: Faculty of Education, Kochi University.

Kenneth S. Olson
ken_olson@sil.org

Emy T. Ballenas
emy_ballenas@sil.org

Nilo M. Borromeo
nilo_borromeo@wycliffe.org